Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 15:59:36
REFERENCE NO: 272

Author Names & Affiliations

- Neil Cobb - Northern Arizona University

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

PO Box: 6077, Flagstaff AZ 86011-6077 Peterson Hall - (Bldg #22)Room #330

Title of Submission

Understanding biodiversity through increased investment in cyberinfrastructure

Abstract (maximum ~200 words).

In the last 15 years we have experienced an explosion in the mobilization of biodiversity data and concomitant development of programs, networks, and workflows that will further increase this rate of biodiversity data creation. These emerging big data consist primarily of records for museum specimens, including occurrence data, genetics, images, and complementary environmental data. Complementing this is a huge amount of information on the physiology, ecology, and evolutionary genetics on hundreds of thousands of species that are available to be integrated with this huge emergence of biodiversity data. Additionally, there are several emerging citizen science efforts that are also generating huge amounts of data (e.g., iNaturalist, National Phenology Network, eBird) that are becoming increasingly more available for use in research. We have reached a point where we lack the ability to integrate all these data-sets together, and we cannot address fundamental questions in ecology and evolutionary biology until we can integrate these disparate sets of data. Also, despite our increase in data generation we need more sophisticated sensor workflows to exponentially increase our ability to mobilize more biodiversity data.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

Biodiversity data has met the qualitative criteria for big data, we have complex and large data sets and very few good mechanisms for integration, archiving, and synthesis. We need both greater investment in centralized and distributed cyberinfrastructure as well as the expertise to sustain cyberinfrastructure-based programs. We also need an increased investment in robotics, computer vision/artificial intelligence, real-time sensor networks. Three examples include 1) a robotic imaging system to process the approximately 500 million specimens in US museums. Such a robotic system would capture all the specimen data as well as produce 360-degree images; 2) We are losing taxonomic experts at an alarming rate, a highly-developed computer vision process would greatly augment human expertise and

provide identification assistance not only for existing museum specimens but for citizen scientist that document the occurrence and phenology of species in the field; 3) although there are some very sophisticated sensor networks that are collecting valuable ecological data on species, these need to become more sophisticated and more accessible to distributed networks of researchers.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

The biodiversity community is in great need for almost all aspects of cyberinfrastructure, including software development, networking, sensor/robotics development, and advanced data infrastructure. This includes support for long-term technical support. For example, we are collaborating on a software management system (Symbiota) for digitizing specimen data and images and annotating records. There are few resources to support programmers to develop this software and virtually no funding to work with other software projects that would benefit the entire biodiversity community.

Consent Statement

- "I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publically available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode)."